Data Collection and Validation

Prepared by:
Charley A. Knoderer
Timothy S. Dye
Sonoma Technology, Inc.
Petaluma, CA
(707) 665-9900

AMS Short Course on the Fundamentals of
Boundary Layer Wind and Temperature Profiling
Using Radar and Acoustic Techniques
Long Beach, CA

February 8, 2003

Data Collection and Validation

Data Collection

- Polling frequency
- Instrument settings
- Data sets
- Data merging
- Site operators

Data Validation

- Definition
- Importance
- QC Levels
- QC Codes
- Common Problems and Examples
- Quality Control Tips
- Rules of Thumb and Validation Tips

Collection–Polling Frequency Data

Polling frequency

- How often?
- Advantages/disadvantages of real-time polling
- Data inspection on a regular basis

Data Collection–Instrument Settings

Instrument settings

- Proper time standard
- Seasonal changes
- Averaging intervals-time/height
- File types-advantages/disadvantages

Data Collection–Data Sets

Primary data sets

- Storage
- Real-time or back up?

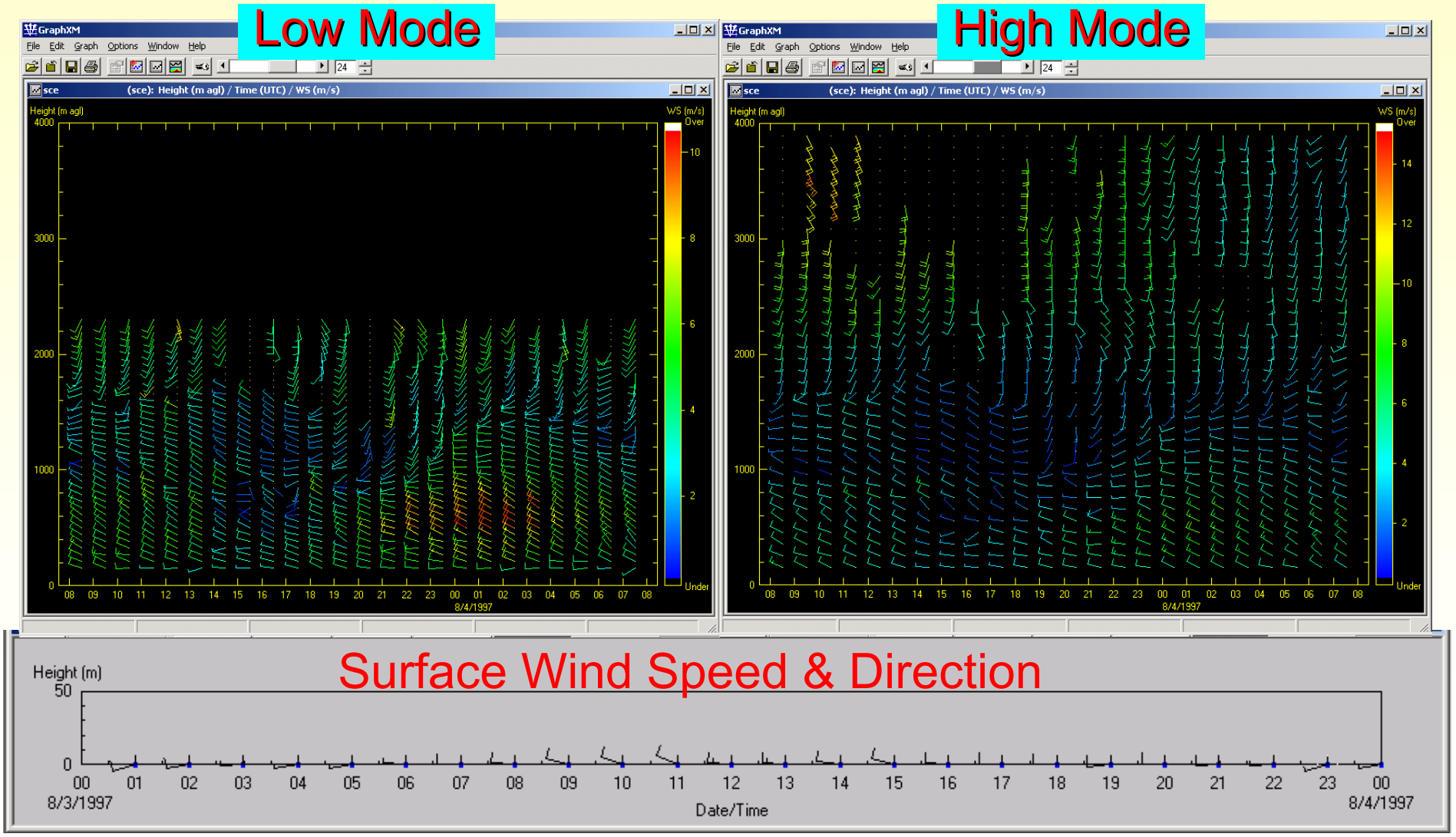
Back up data sets

- Importance
- Off-site storage

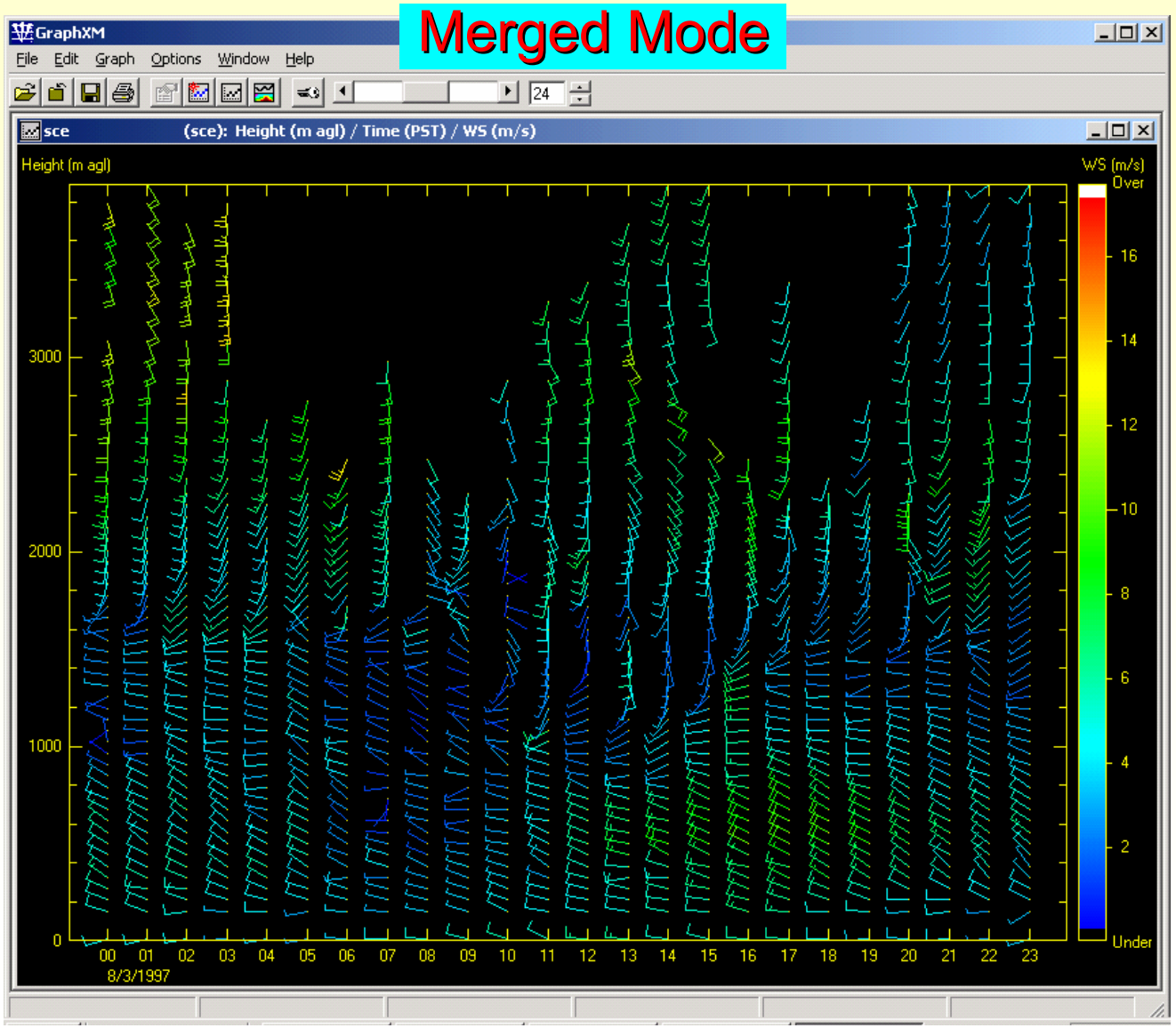
Data Collection–Data Merging (1 of 3)

- Modes
  - Low
  - High
- Procedure for merging
- Merge surface data with upper-air data

Data Collection–Data Merging (2 of 3)



Data Collection–Data Merging (3 of 3)



Data Collection–Site Operators

- Training
- Site visit frequency
- Tasks and responsibilities
  - Record visit in site log
  - Complete maintenance checklist
  - Inspect RASS audio sources
  - Stop radar wind profiler
  - Archive radar wind profiler's moments and consensus data
  - Check computer clock
  - Inspect radar wind profiler
  - Restart radar wind profiler
  - Secure shelter
  - Submit maintenance checklist and site log

Data Validation (1 of 4)

Definition

Importance

QC levels

QC codes

Common problems and examples

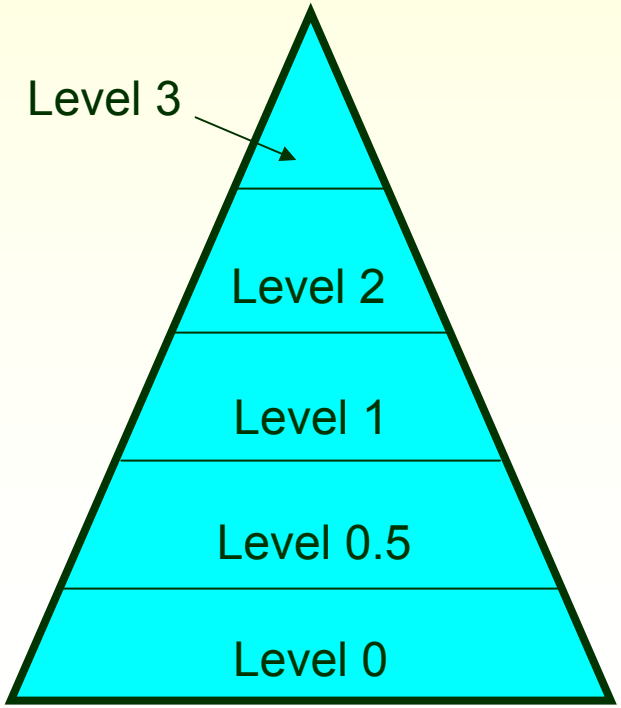
Eight quality control tips

Rules of thumb and validation tips

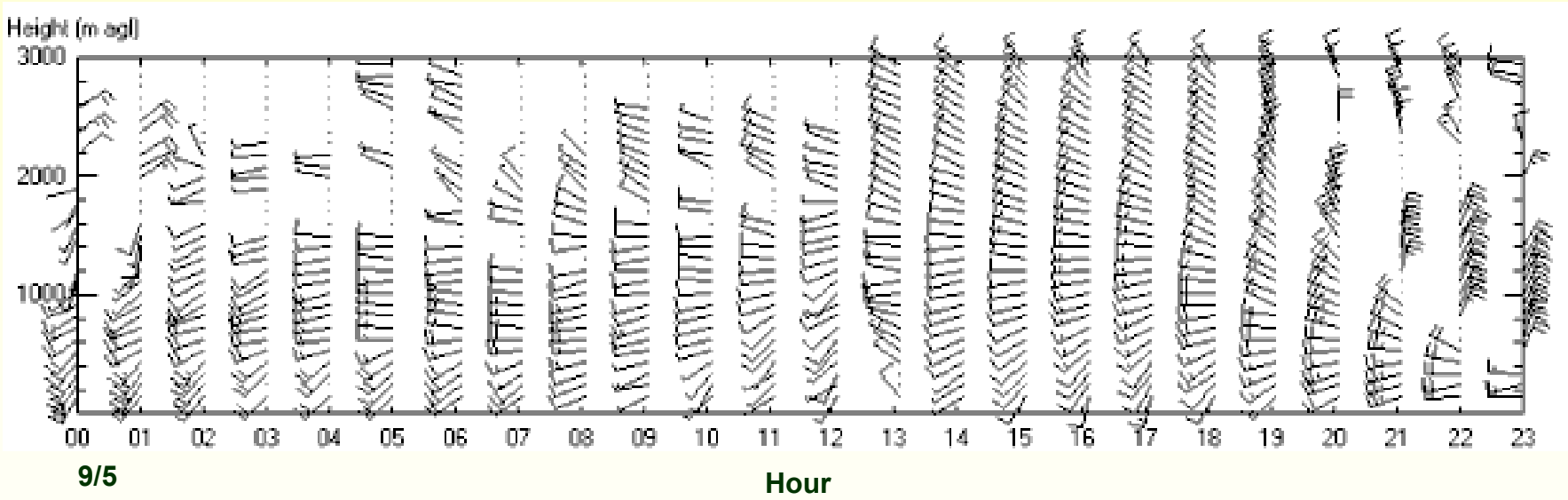
Data Validation (2 of 4)

Definition: the process of identifying inconsistent observations (outliers) and assigning QC codes to each data point to indicate its validity (Watson et al., 1998).

Outliers: Data that are spatially, temporally, or physically inconsistent.

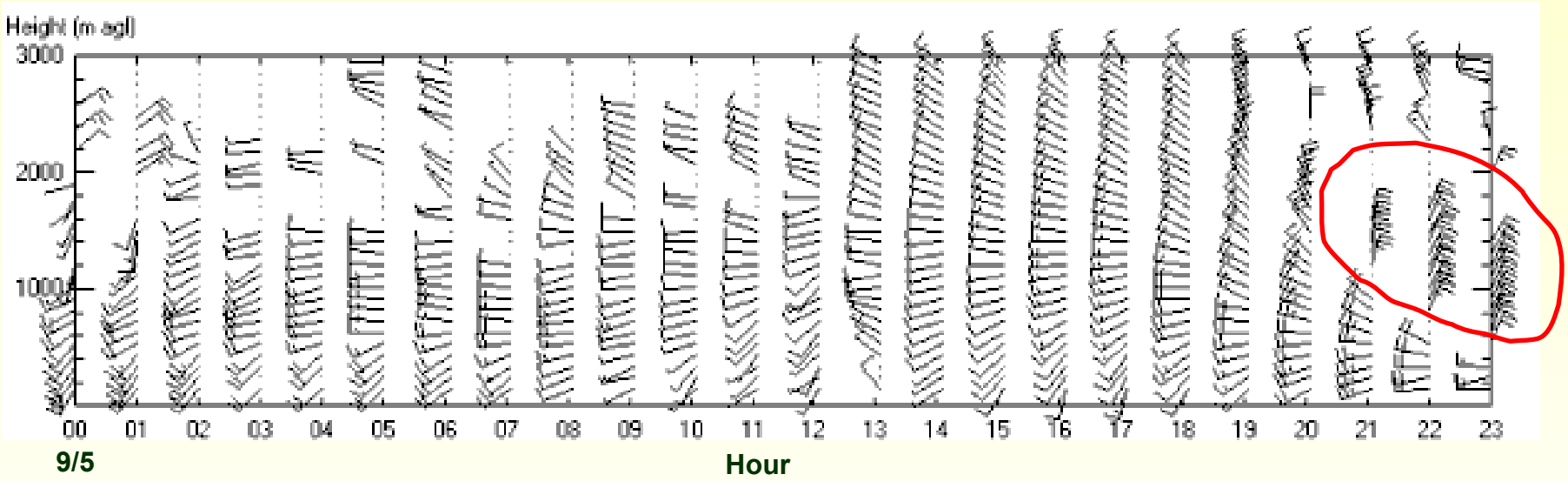


Why Validate the Data? (3 of 4)



Data Validation (4 of 4)

"Good looking" data can be invalid, bad, inconsistent, etc!



This shows a time-series plot of wind speed and direction at various altitudes for a 24-hour period with bird contamination. The orientation of the barb indicates wind direction (barb or flag facing up = wind from north). A larger number of tails on the barbs indicates increasing wind strength.

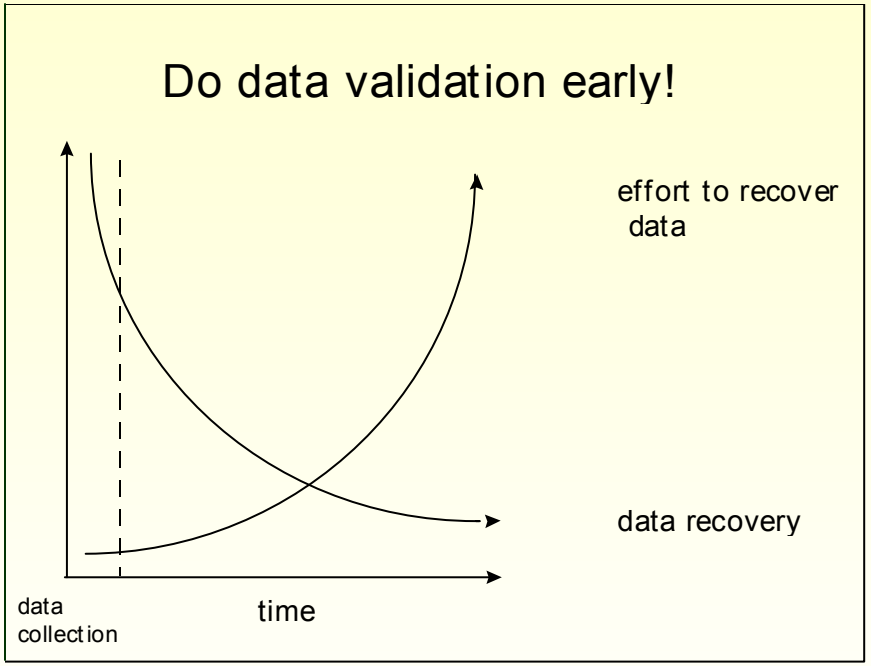
The northerly winds from 2100 and 2300 EST between 500 and 2000 m above ground level (agl) were actually caused by the radar measuring the motion of birds migrating south, instead of the northwesterly atmospheric winds. Birds act as large radar "targets," so that signals from birds overwhelm the weaker atmospheric signals.

Birds generally migrate year-round along preferred flyways, with the peak migrations occurring at night during the spring and fall months (Gauthreaux, 1991). Additional information about bird contamination of radar wind profiler data can be found in Wilczak et al. (1995).

Data Validation–Importance

Data validation is critical because serious errors in data analysis and modeling results can be caused by erroneous individual data values.

Timely data validation is required to minimize the generation of additional data that may be invalid or suspect and to maximize the recoverable data.



Data Validation–QC Levels (1 of 3)

Level 0 Data Validation: Routine checks are made during the initial data processing and generation of data, including proper data file identification, review of unusual events, review of field data sheets and result reports, and instrument performance checks.

- Verify computer file entries against data sheets.
- Flag samples when significant deviations from measurement assumptions have occurred.
- Eliminate values for measurements that are known to be invalid because of instrument malfunctions.
- Replace data from a backup data acquisition system in the event of failure of the primary system.
- Adjust measurement values of quantifiable calibration or interference bias.
- Document the changes made to the data.

Data Validation–QC Levels (2 of 3)

Level 0.5 Data Validation: Automatic (objective) checks are applied to the data to identify outliers.

Types of checks

- Range
- Rate of change
- Pattern recognitions (Webber-Wuertz)

Level I Data Validation: Manual review of data for internal consistency to identify values that appear atypical when compared to values for the entire data set and to the reviewer's knowledge of expected meteorological conditions.

- Compare data collected from nearby sites at similar heights and times.
- Compare data to surface meteorology.

Data Validation–QC Levels (3 of 3)

Level II Data Validation: Comparison of the current data set with historical data by an experienced meteorologist to verify consistency over time. This level is often part of the data interpretation or analysis process.

- Compare data to data collected from other instruments (nearby profilers, rawinsondes) or upper-air maps.

Level III Data Validation: Occurs when the data are used during modeling and analysis efforts, for example, when inconsistencies in analysis and modeling results are found to be caused by measurement errors.

Data Validation–QC Codes

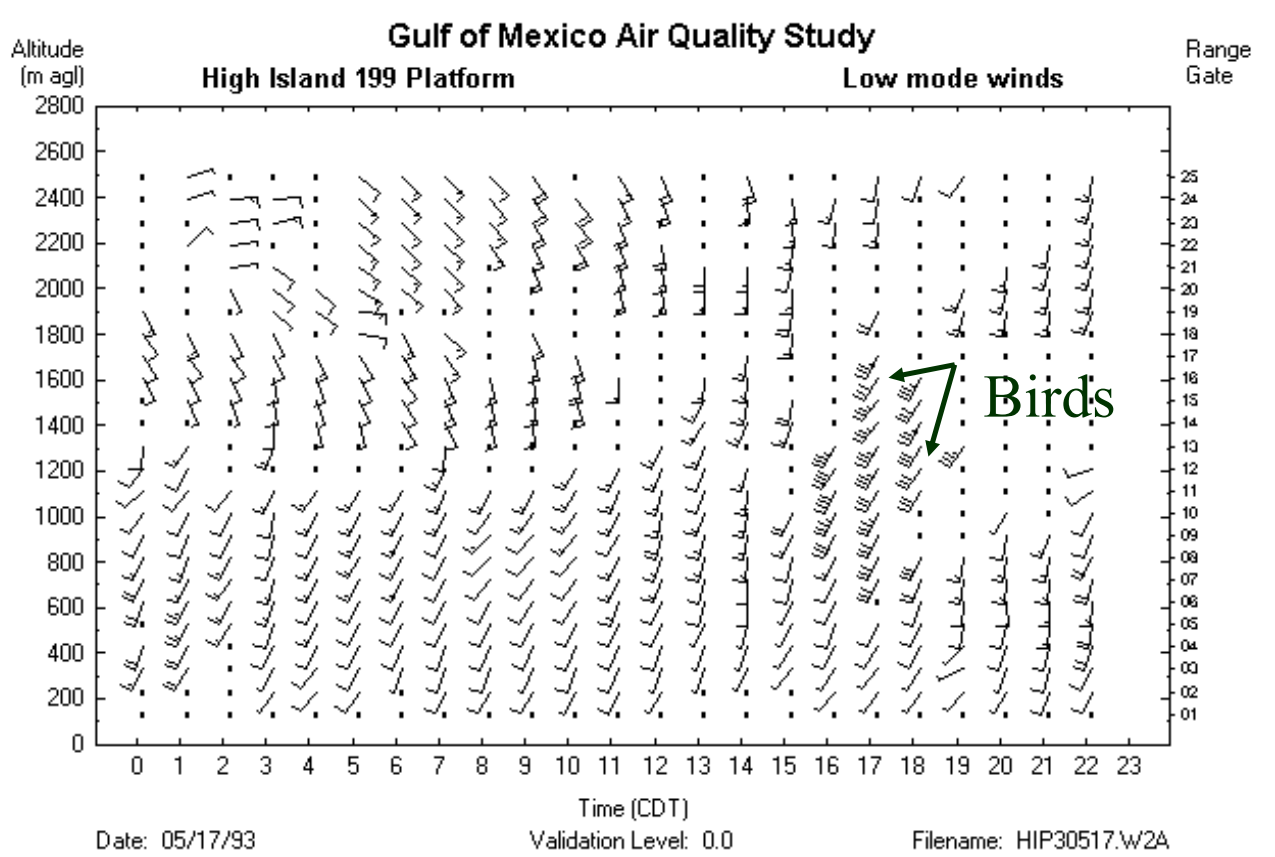
QC Code	QC Code Name	Definition
0	Valid	Observations that were judged accurate within the performance limits of the instrument.
1	Estimated	Observations that required additional processing because the original values were suspect, invalid, or missing. Estimated data may be computed from patterns or trends in the data (e.g., via interpolation), or they may be based on the meteorological judgment of the reviewer.
2	Calibration applied	Observations that were corrected using a known, measured quantity (e.g., instrument offsets measured during audits).
3	Unassigned	Reserved for future use.
4	Unassigned	Reserved for future use.
5	Unassigned	Reserved for future use.
6	Failed automatic QC check	Observations that were flagged with this QC code did not pass screening criteria set in automatic QC software.
7	Suspect	Observations that, in the judgment of the reviewer, were in error because their values violated reasonable physical criteria or did not exhibit reasonable consistency, but a specific cause of the problem was not identified (e.g., excessive wind shear in an adiabatic boundary layer). Additional review using other, independent data sets (Level 2 validation) should be performed to determine the final validity of suspect observations.
8	Invalid	Observations that were judged inaccurate or in error, and the cause of the inaccuracy or error was known (e.g., winds contaminated by ground clutter or a temperature lapse rate that exceeded the autoconvective lapse rate). Besides the QC flag signifying invalid data, the data values themselves should be assigned invalid indicators.
9	Missing	Observations that were not collected.

Data Validation–Common Problems (1 of 7)

- Interference from migrating birds
- Precipitation interference
- Ground clutter interference
- Velocity folding or aliasing
- Vertical velocity correction (RASS)
- Cold bias (RASS)

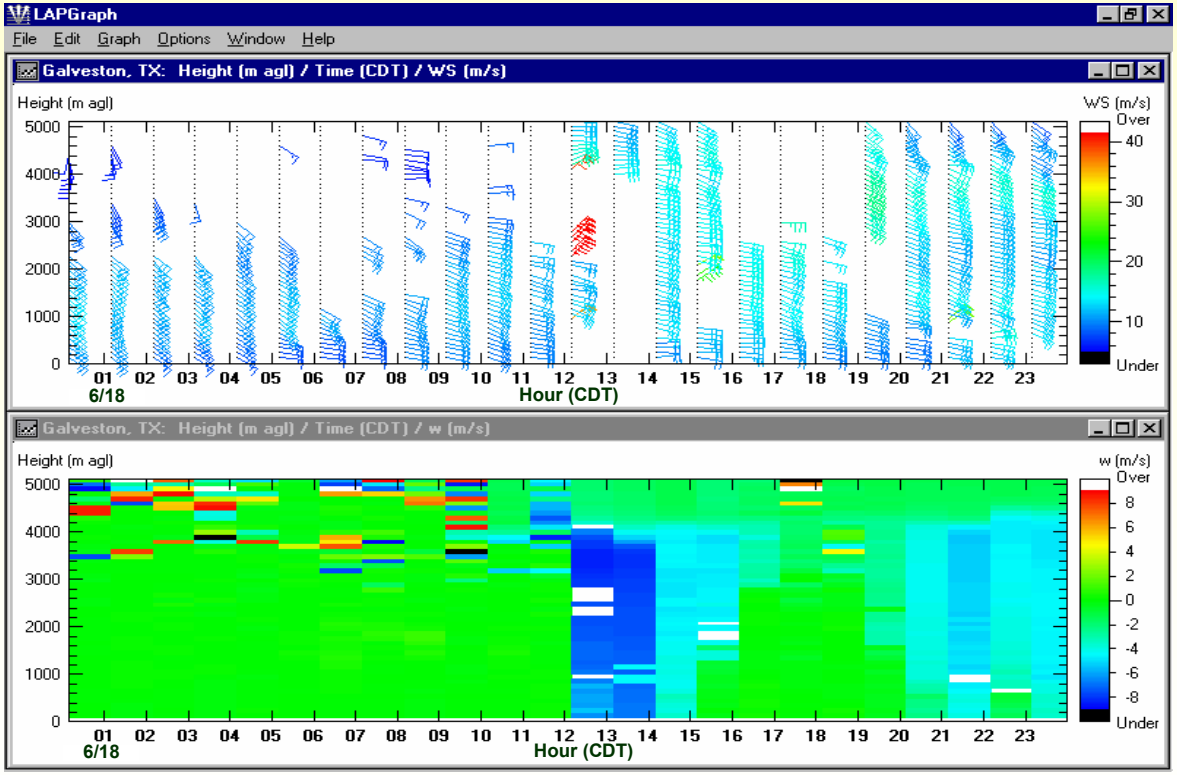
Data Validation–Common Problems (2 of 7)

Interference from migrating birds: Birds act as large radar targets so that signals from birds overwhelm the weaker atmospheric signals. *This can produce biases in the wind speed and direction measurements.*



Data Validation–Common Problems (3 of 7)

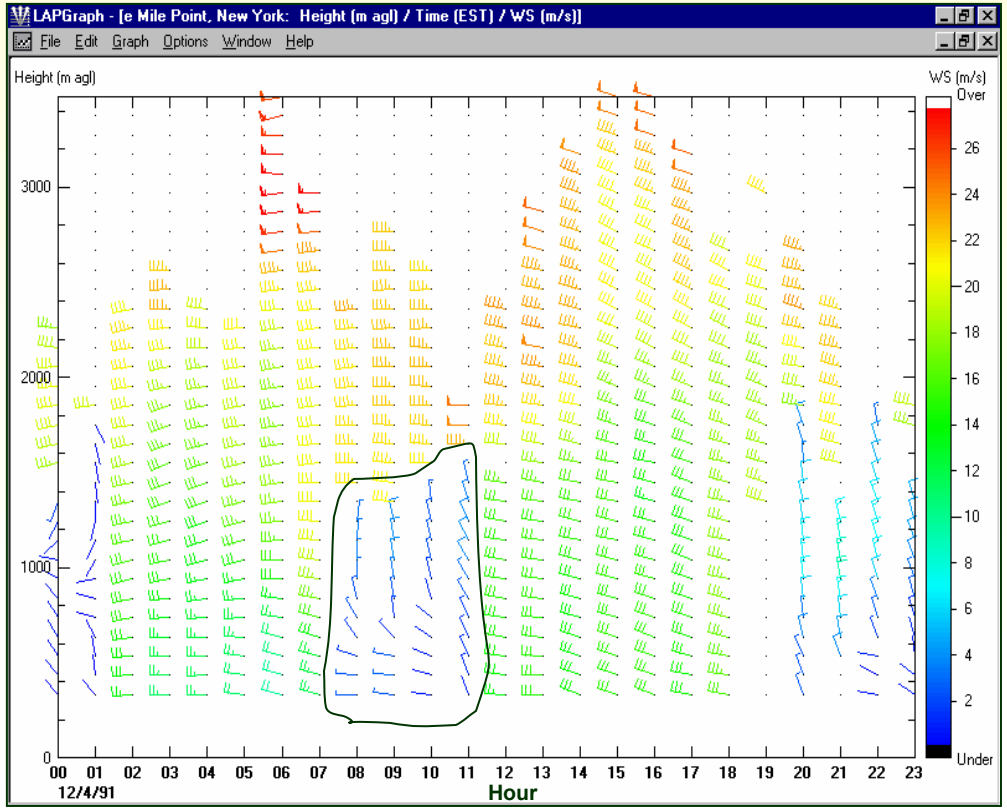
Precipitation interference: During precipitation, the profiler measures the fall speed of rain drops or snowflakes.



Precipitation Interference. Example of precipitation interference in radar profiler data collected at Galveston, TX, on June 18, 1993. Missing wind data at 1200, 1300, and 1900 CDT were caused by precipitation. In this example, the profiler had difficulty measuring winds between 1200 and 1500 CDT during convective (i.e., highly variable) rain. When the rain was stratiform in nature, the profiler was capable of measuring winds.

Data Validation–Common Problems (4 of 7)

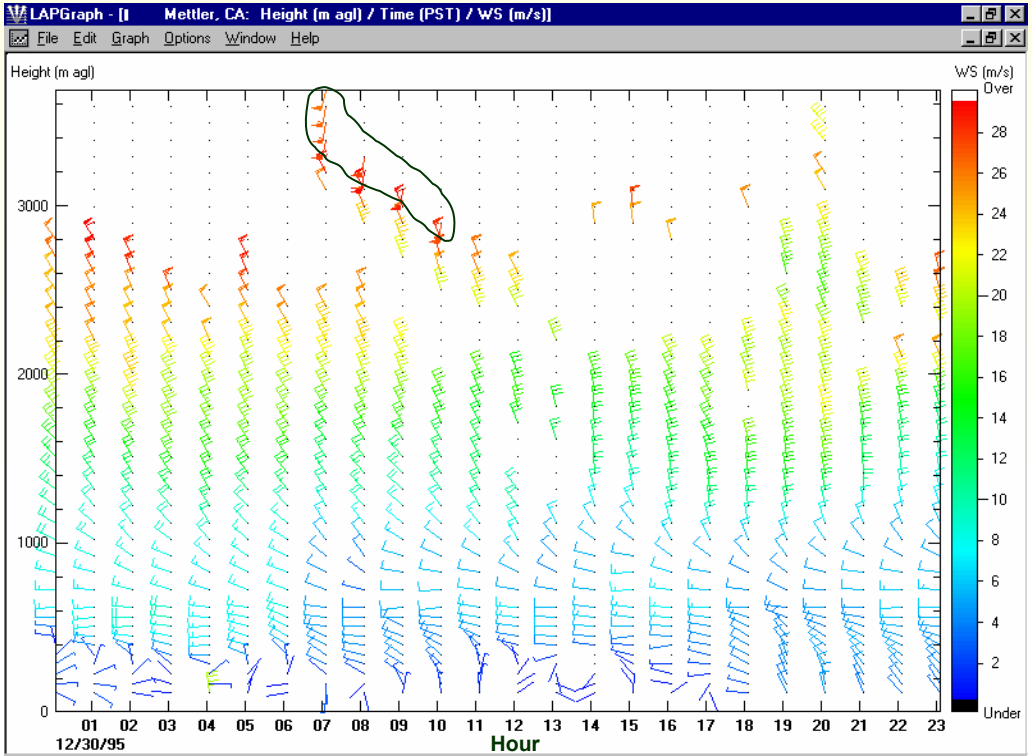
Ground clutter: Ground clutter occurs when a transmitted signal is reflected off of objects such as trees, power lines, or buildings instead of the atmosphere. *Data contaminated by ground clutter can be detected as a wind shift or a decrease in wind speed at affected altitudes.*



Ground Clutter. Example of ground clutter interference at a radar profiler site in Nine Mile Point, NY. At this site, the beams are oriented almost north and east. The calm winds are a result of ground clutter interfering with both beams. The weak northerly winds are a result of ground clutter interference affecting only the east beam.

Data Validation–Common Problems (5 of 7)

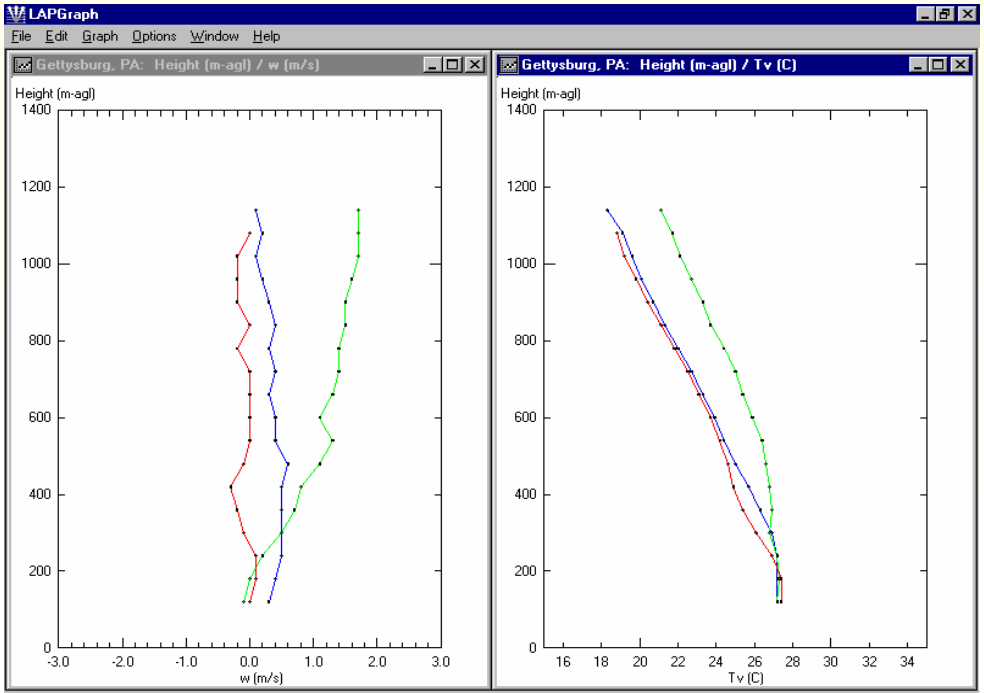
Velocity folding or aliasing: Velocity folding or aliasing occurs when the magnitude of the radial component of the true air velocity exceeds the maximum velocity that the instrument is capable of measuring. *Folding occurs during very strong winds (>20 m/s) and can be identified and flagged by automatic screening checks* (Miller et al., 1994).



Velocity folding or aliasing. In this example, the southerly winds between 0700 and 1000 PST around 2800 to 4000 m agl were caused when the winds exceeded the Nyquist velocity.

Data Validation–Common Problems (6 of 7)

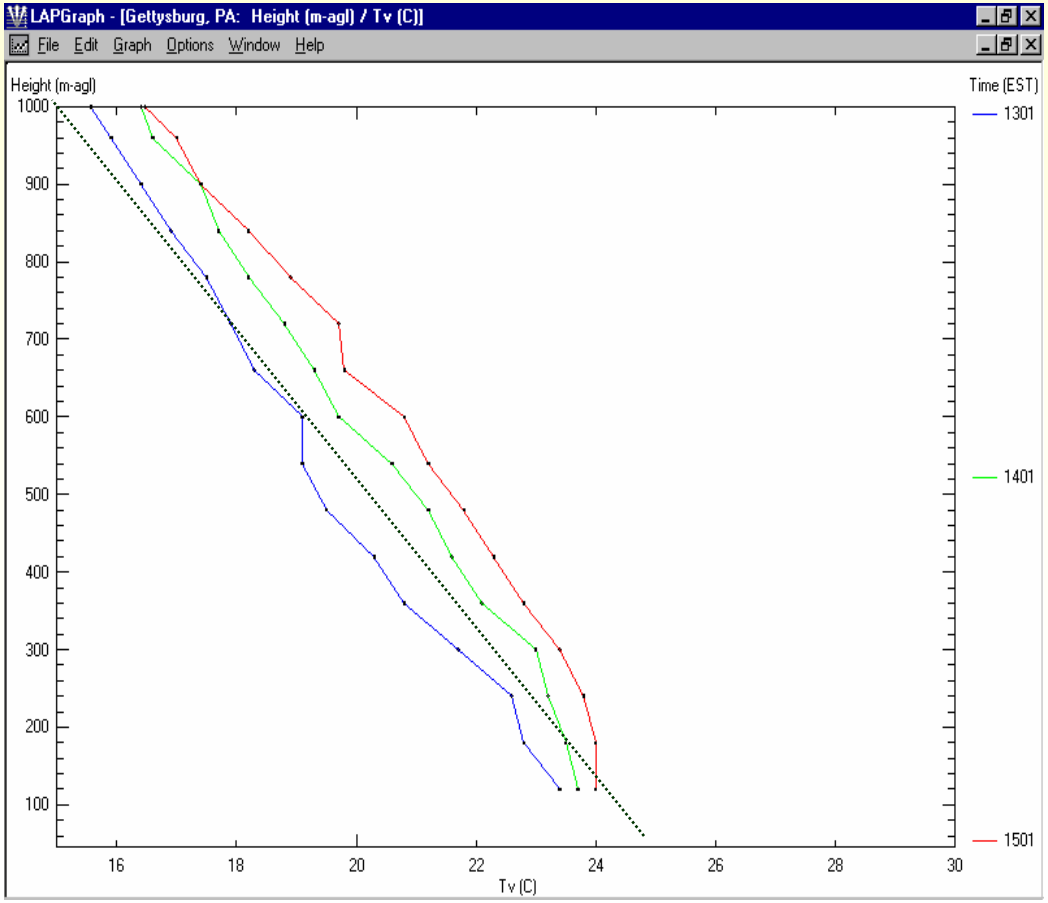
Vertical velocity correction (RASS): Vertical motions can affect the RASS virtual temperature measurements. Virtual temperature is determined by measuring the vertical speed of an upward-propagating sound pulse, which is a combination of the acoustic velocity and the atmospheric vertical velocity. *If the atmospheric vertical velocity is non-zero and no correction is made for the vertical motion, it will bias the temperature measurement.*



Vertical Velocity Interference. Example of vertical velocity (w) and its affect on the uncorrected RASS virtual temperature (T_v) measurements. Notice that when the vertical velocity is near zero, the temperature profile is generally adiabatic and meteorologically consistent. However, when the vertical velocity is non-zero, it alters the acoustic velocity, which if not corrected introduces a bias in the virtual temperature. In this case, a 1 to 2 m/s updraft altered the virtual temperature by 1.5 to 2.0°C and produced an artificial warming.

Data Validation–Common Problems (7 of 7)

Cold bias (RASS): *Under certain conditions (possibly associated with site selection issues), RASS observations may exhibit a bias of -1 °C or so.*



Cold bias. Recent intercomparisons between RASS systems and rawinsondes have shown a bias in the first couple of sampling altitudes (Riddle et al.,1995). The RASS virtual temperatures are often slightly cooler (-0.5 to -1.0°C) than the reference rawinsonde data. Work is currently underway to address this issue.

Data Validation–Eight Quality Control Tips (1 of 2)

1. Display upper-air data using GraphXM, LapGraph, or a similar editing program.
2. Review 24 hours at a time with scales set to automatic. When quality controlling vertical profiles, such as virtual temperature, review three hours at a time, scrolling through each profile one hour at a time.
3. Make sure that no data exist above plotting altitude.
4. Make sure auto-range is selected to obtain the maximum height of available data.

Data Validation–Eight Quality Control Tips (2 of 2)

5. When quality controlling the low and high modes separately, look at them together to make sure they are consistent with one another.
6. Try to set QC codes to valid or invalid.
7. Questionable data may be compared with model winds or nearby upper-air data.
8. At night, consider shear layers near the surface carefully. Wind direction may be highly variable and wind speeds may be very weak.

Data Validation–Rules of Thumb and Validation Tips (1 of 2)

Criteria	Action	Cause
Less than 50% of wind data	Invalidate	Computer Reboots or Power Outages
Sparse, isolated, and inconsistent data aloft	Invalidate	Likely due to low signal to noise ratio (SNR)
High, inconsistent winds in first one through five range gates	Invalidate	Ringing in receiver
Inconsistent at top of profile	Invalidate	Low signal to noise ratio (SNR)
Rainfall-constant	Validate	
Rainfall-convective	Invalidate	
Superadiabatic in first few range gates	Validate	
Aloft inversion	Validate if time consistent	
Temperature Profile shifted (inconsistent with other profiles)	Invalidate	Vertical Velocity contamination ($\pm$ 1 m/s = $\pm$ 1.7 °C)
Convective site	Use corrected virtual temperature	
Non-convective site	Use uncorrected virtual temperature	

Data Validation–Rules of Thumb and Validation Tips (2 of 2)

Expected wind direction variability with adjacent (time & height) winds.

Wind Speed	Wind Direction Variability	Comments
0-2.5 m/s	180°	
2.5-5.0 m/s	90°	
5.0-10.0 m/s	30°	Except in the case of shear or an inversion. Compare to temperature profile.
Greater than 10.0 m/s	20°	

Data Collection and Validation–Summary (1 of 2)

- Data collection should be done on a regular basis.
- Proper time standards should be applied.
- Seasonal changes need to be accounted for when setting the profiler configuration.
- Different file types serve different purposes.
  - For example, real-time data are not efficiently transferred over phone lines using database files.
- Back-up data sets should be stored in more than one location.

Data Collection and Validation–Summary (2 of 2)

- Data validation is the process of identifying inconsistent observations and assigning QC codes to each data point to indicate its validity.
- Data validation is important because serious errors in data analysis and modeling results can be caused by erroneous individual data values.
- QC levels identify the “level” of effort involved in quality controlling the data.
- QC codes are assigned to individual data points to indicate their validity.
- Several sources of interference exist:
  - Migrating birds
  - Precipitation
  - Ground clutter
  - Velocity folding
  - Vertical velocity correction (RASS)
  - Cold bias (RASS)

References

Gauthreaux S.A. (1991) The flight behavior of migrating birds in changing wind fields: Radar and visual analyses. *American Zoologist* **31**(1), (in press).

Miller P.A., Schlatter T.W., van de Kamp D.W., Barth M.F., and Weber B.L. (1994) An unfolding algorithm for profiler winds. *J. Atmos. Ocean. Technol*. **11**, 32-41.

Riddle A.C., Angevine W.M., Ecklund W.L., Miller E.R., Parsons D.B., Carter D.A., and Gage K.S. (1995) Wind-profiler performance and balloon observations compared at ISS sites during TOGA COARE. *Contrib. Atmos. Phy.*, (accepted for publication).

Wilczak J.M., Strauch R.G., Weber B.L., Merritt D.A., Ralph F.M., Jordan J.R., Wolfe D.E., Lewis L.K., Wuertz D.B., Gaynor J.E., McLaughlin S., Rogers R., Riddle A., and Dye T. (1995) Contamination of wind profiler data by migrating birds: characteristics of corrupted data and potential solutions. *J. Atmos. Ocean. Technol.* **12**, 449-467.

Watson J.G., DuBois D.W., DeMandel R., Kaduwela A., Magliano K., McDade C., Mueller P.K., Ranzieri A., Roth P.M., and Tanrikulu S. (1998) Aerometric monitoring program plan for the California Regional $PM_{10}/PM_{2.5}$ Air Quality Study. Prepared for California Regional $PM_{10}/PM_{2.5}$ Air Quality Study Technical Committee, California Air Resources Board, Sacramento, CA by Desert Research Institute, Reno, NV, DRI Document No. 9801.1D5, December.